DOCUMENT RESUME

ED 417 198 TM 028 139

AUTHOR	McLaughlin, Don
TITLE	Can State Assessment Data Be Used To Reduce State NAEP Sample Sizes? Working Paper Series.
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-WP-97-29
PUB DATE	1997-10-00
NOTE	31p.
AVAILABLE FROM	U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Avenue, N.W., Room 400, Washington, DC 20208-5654.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Achievement; Comparative Analysis; Correlation; Costs; Data Collection; Educational Assessment; Elementary Secondary Education; Longitudinal Studies; National Surveys; Research Methodology; *Sample Size; *State Programs; *Test Results; *Testing Programs
IDENTIFIERS	*Linkage; *National Assessment of Educational Progress; State Assessment of Educational Progress

ABSTRACT

In the 1970s and 1980s, the National Assessment of Educational Progress (NAEP) built a longitudinal record as the Nation's Report Card based on periodic brief assessments of a modest but representative sample of the nation's students. In 1990, data collection was expanded from a sample of 10,000 students in each grade and subject area to 100,000 to provide the basis for state-by-state comparisons. This expansion brought about an large increase in data collection costs. There is a frustrating conflict between the need for precise estimates of educational achievement and the cost of obtaining these estimates. This study considers using state assessments to supplement, or reduce, NAEP samples. The relationships between using state assessment scores and reducing sample size are demonstrated mathematically. Implementing this approach would require step-by-step planning and implementation. While the linkage of student-level assessment scores would be ideal, linkages based on school-level summary statistics appear to be sufficient when the correlations between tests is high. Many states appear to have tests with such correlations to the NAEP. The cost of linking procedures required for implementing the sample size reductions is in the range of $5,000 to $10,000 per state, which is a small percentage of the cost of the administration of State NAEP in an additional 50 schools in the state. Many states have difficulty recruiting schools for the NAEP and would welcome this initiative. These analyses suggest that several states could be involved in sample size reduction, possibly as early as 1998. (Contains one table and one figure.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

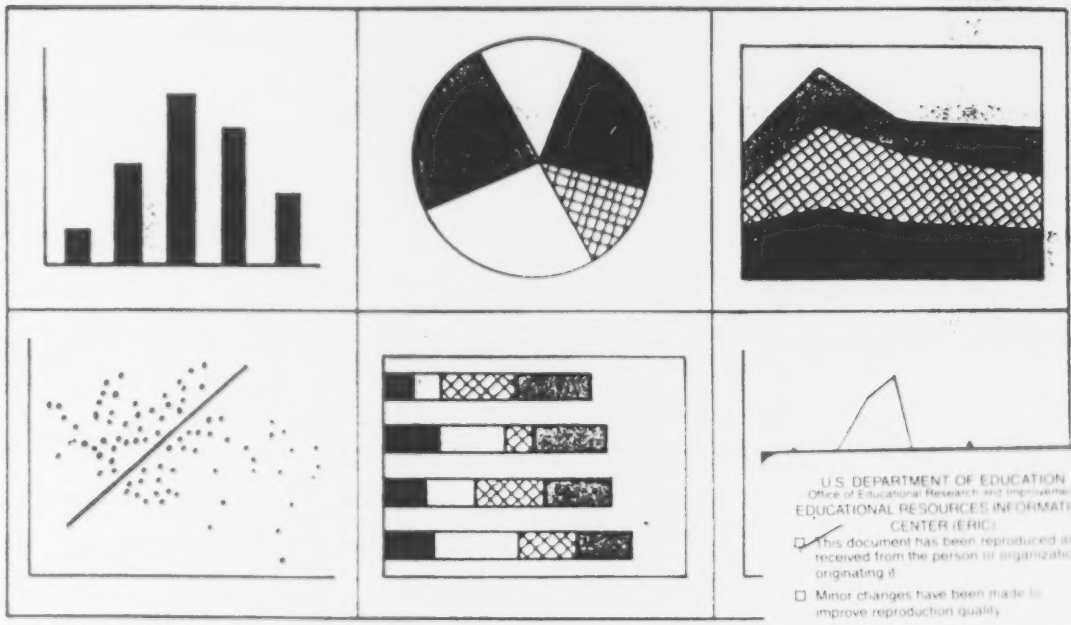
ED 417 198

NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?

Working Paper No. 97-29 October 1997



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM028139

U.S. Department of Education
Office of Educational Research and Improvement

BEST COPY AVAILABLE

Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?

Working Paper No. 97-29 October 1997

Contact: Steven Gorman
Assessment Group
(202) 219-1937
e-mail: steven_gorman@ed.gov

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Ricky T. Takai
Acting Assistant Secretary

National Center for Education Statistics
Pascal D. Forgione, Jr.
Commissioner

Assessment Group
Gary W. Phillips
Associate Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue, NW
Washington, DC 20208

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?*, Working Paper No. 97-29, by Don McLaughlin. Project Officer, Steven Gorman. Washington, D.C.: 1997.

October 1997

Foreword

Each year a large number of written documents are generated by NCES staff and individuals commissioned by NCES which provide preliminary analyses of survey results and address technical, methodological, and evaluation issues. Even though they are not formally published, these documents reflect a tremendous amount of unique expertise, knowledge, and experience.

The ***Working Paper Series*** was created in order to preserve the information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series. Consequently, we encourage users of the series to consult the individual authors for citations.

To receive information about submitting manuscripts or obtaining copies of the series, please contact Ruth R. Harris at (202) 219-1831 or U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Ave., N.W., Room 400, Washington, D.C. 20208-5654.

Samuel S. Peng
Acting Director
Statistical Standards and Services Group

Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?

Prepared by:

Don McLaughlin

Prepared for:

U.S. Department of Education
Office of Educational Research and Development
National Center for Education Statistics

October 1997

Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?

Don McLaughlin
July, 1997

Background

Initiated in the late 1960s, the National Assessment of Educational Progress (NAEP) is the outstanding example of this nation's attempts to improve the education of its children by bringing information about children's achievement to the awareness of the electorate. Throughout the 1970s and 1980s, NAEP built a longitudinal record as the Nation's Report Card, based on periodic brief assessments of a modest, but representative sample of the nation's 9, 13, and 17 year-olds (more recently, 4th, 8th, and 12th graders) in a variety of areas. Characterized by state-of-the-art assessment design and data analysis, NAEP's reputation has grown to merit consideration as the "gold standard" of educational testing, a model for other testing programs.

Data collection for NAEP was expanded in 1990 from a sample of 10,000 students in each grade and subject area to 100,000, to provide the basis for state-by-state comparisons. That expansion brought with it an increase in data collection costs by an order of magnitude -- costs borne not only by the U.S. Department of Education but also by participating states, through the in-kind efforts they contributed to support the sampling and to administer the tests. In addition to the millions of dollars spent by the federal government, McLaughlin et al. (1993) estimated that each state contributed more than $50,000 in effort to participate in one grade and subject area.[1]

The Problem

There is a frustrating conflict between the need for precise estimates of educational achievement and the cost of obtaining those estimates. Early in the history of NAEP, the conflict was over testing time. The original designers of NAEP initiated the first innovation to reduce testing burden: matrix sampling. In the 1980s, the Educational Testing Service developed additional innovations, such as booklet BIB spiraling to enable reporting of valid achievement levels in the context of limited testing.

---

[1] This was a background study to the National Academy of Education's *Evaluation of the 1990 Trial State Assessment*.

With the expansion in 1990, the size of the sample of schools that must participate in each state has become the focus of the conflict. State education agency staff have struggled, and in some cases failed, to obtain the participation of a sufficient percentage of their NAEP-sampled schools. If the sample size could be reduced, participation would not only be easier for states but also less costly to the federal government, which has paid for the printing and processing of the tests and the training of the test administrators.

At the same time, schoolteachers and principals continue to decry the vast amount of time spent on testing. NAEP is particularly vulnerable to these cries because, unlike state- and district-run standardized testing and assessment programs, NAEP is not designed to give diagnostic information back to schools, teachers, and parents. At the local level, when asked to participate, it is not surprising that superintendents wonder whether NAEP isn't redundant, although they can see the value of a national indicator.

A Possible Solution

In most states, as catalogued by the Council of Chief State School Officers, state assessments are now being administered. States are very concerned about the educational achievement of their schoolchildren, many have implemented far-reaching school reform efforts, and they want accountability. Therefore, they are calling on test publishers to develop ever more valid tests of what children should know and be able to do. They are administering these tests to all the children in their states. And they are asking, how do their tests relate to the "gold standard," to NAEP?

Why not incorporate linkage of state assessments to NAEP as a part of NAEP? Not only would this be of value to the states, but if it should be the case that the state assessments ARE highly correlated with NAEP, and current work by the author suggests that they are, then the sample needed for NAEP can be substantially reduced in those states. To the extent that state assessment information accounts for variation in NAEP scores, the sampling variability in NAEP estimates for states can be virtually eliminated, because state assessments are generally administered in all schools in a state. Measurement error remains, but the elimination of a great deal of the sampling error means that an equal precision can be obtained with a much smaller sample of schools (e.g., half as many).

This is almost "getting something for nothing," and as such, it should evoke caution. The purpose of this paper is to lay out the issues to be addressed in deciding whether to use state assessments to supplement NAEP samples, to provide a framework for a serious evaluation of this innovation, and to demonstrate mathematically the relations between using state assessment scores and reducing sample size.

Framework for a Solution

The proposed solution is not a simplification, and there is a need to spell out in detail

what it takes to implement the solution -- that is, to incorporate state assessment information in NAEP estimation and reduce the NAEP sample thereby. As a first step, it will help to lay out the steps in the process that would occur once the innovation is in place. (If those steps prove hypothetically feasible, the next step will be to lay out the steps needed to put the innovation in place.) Implementing the solution in the context of a particular NAEP assessment requires five steps.

1. *An overall framework must be developed, indicating which states might reasonably contribute state assessment information.*

NAEP must start with a knowledge of which states have the potential for sample size reduction in a particular year. In most cases, prior information on the correlation of the state assessment with NAEP, at least at the school level, will be available; but when it is not available, as when a state implements a new state assessment, a pilot test of the linkage may be warranted. The relations between this step and the national field tests proposed in the redesigned NAEP must be explored. Although this innovation can be implemented independently in each state, the savings to the federal government will be proportional to the number of states in which it is implemented, so accurate decisions on which states to include in the sample size reduction are important.

2. *Arrangements for cooperation with each state must be made.*

One cost of testing in fewer schools is the addition of a few data management procedures to enable linkage of NAEP with state assessment scores. The cost of these procedures will be minimized if they can be carried out centrally in each State Education Agency office. Whether that will be possible, as it was in the study of NAEP-to-state assessment linkage carried out by the author, or whether it must involve effort at local schools, depends on the nature of the assessment data collection system in the state. It is probable, however, that in the future nearly every state with a viable state assessment will have a central data management facility.

3. *A NAEP sampling plan must be developed and implemented in each state.*

The NAEP sampling plan would not need to be different from the present plan, except for a smaller sample size, estimated to be sufficient given prior information about the probable correlation of state assessments with NAEP. It should be pointed out that the innovation described here could complement other innovations to reduce sample size without using state assessment data.

4. *Procedures must be developed and implemented for maintaining confidentiality while linking state assessment scores to NAEP.*

Parents are concerned that their children's test scores and answers to background questions not be made public. NAEP takes pride in the care with which they maintain the

confidentiality of NAEP information, as state assessment directors take pride in the security of their assessment data. Maintaining this confidentiality, it is still possible to link state assessment scores to NAEP records anonymously for the purposes of developing linkage formulas and population estimates. The author has demonstrated this in four states as a part of study of state assessment-to-NAEP linkage possibilities. In cases in which state assessment data are available during the NAEP student sampling process, it is possible to implement the linkage merely by inserting state assessment scores on the NAEP administration schedule. In the more common case in which state assessments are administered two or three months after NAEP, an additional step of creating a secure linkage code, preferably at the time of student sampling, is necessary.

One possibility for circumventing the student confidentiality issue is to base the linkage solely on school-level summary information about performance on the state assessment. Although this is conceptually valid, there is a question as to whether the requirements for precision of NAEP reporting could be met using a linkage based on school-level information.

5. *Analytic procedures must be implemented for using state assessment information appropriately in NAEP estimation.*

Using information on students who participate in NAEP in a state, a function estimating the distribution of NAEP scores for students who do not participate in NAEP, but who have state assessment scores, can be generated. That function may include additional demographic information on individuals and schools. For example, to the extent that gender and race differences on a state assessment are not the same as on NAEP, adjustments for these factors must be included in the estimation. Implementing the estimation must take into account factors that go beyond statistics, however. The logistics of transferring scores from either the State Education Agency or the agency's testing contractor to those responsible for NAEP estimation must be carefully planned so that it becomes impossible for State NAEP reporting at the national level to be held up by unexpected state assessment problems in one or more states.

Implementation Steps for a Solution

Step 1. Invitation

The first step is the decision concerning which states are to be invited to participate in the sample size reduction option. The requirement is that the state have a state assessment that is likely to be linkable to NAEP. The state assessment must be on the same topic as the NAEP assessment (e.g., mathematics) and must be administered to all public school students in grade 4 or 8 in the state, with possible exclusions similar to exclusions from NAEP for students with limited English proficiency or disabilities requiring individual educational programs. There must be some evidence, for example from a previous state NAEP, that performance on the state assessment is strongly correlated with NAEP performance. For example, the requirement might be that the school-level correlation be greater than .75. In 1993-94, this was satisfied for 4th grade reading in at least 10 states, possibly many more. Generally, it is straightforward for

NCES to compute this correlation, given state assessment school means for the most recent matching administration of state NAEP in the state.

The invitation will spell out the specific processes and outcomes of participation in the sample size reduction option, to facilitate a prompt decision by the state as to whether or not to accept the invitation. That spelling out might be based on an adaptation of the following description of the remaining four steps.

Step 2. Sample Selection.

The sample in a state for each grade and assessment topic for which sample reduction is planned will consist of two matched samples of 50 schools. One sample of 50 schools will be the NAEP administration sample, and the other will be a state assessment verification sample. No NAEP data collection will be carried out in the state assessment verification sample, but after both NAEP and state assessment data are collected, simple comparisons of state assessment results between the two samples will verify that the administration of the state assessment in the NAEP sample was no different from the administration of the state assessment in other schools in the state. The comparisons will include exclusion rates, absence rates, and the distribution of performance on the state assessment.

The rationale for the state assessment verification sample is that the only major threats to the validity of the linkage, if the tests are found to be correlated, would be differences in the way in which the state assessment is administered in the NAEP, as compared to other schools in the state. For example, if the LEP exclusion percentage for the state assessment is different from that for NAEP, that would not be a threat to the linkage unless the state assessment LEP exclusion procedures, and therefore results, were different in the NAEP schools than elsewhere. (If the LEP exclusion rates were substantially different between NAEP and the state assessment, however, that factor would need to be included in the linkage analysis to ensure that reporting for subgroups with high percentages of LEP students in the state would not be biased.)

The identity of the schools in the state assessment verification sample do not need to be disclosed until the time at which the comparisons were carried out, after test administration. Therefore, from the perspective of the state test administration, any school might be included in the verification.

Step 3. Linkage data file production.

Two alternatives are being considered: (1) linkage at the student level, and (2) linkage at the school level. Originally, only the former was under consideration, because both the precision and the credibility of the linkage are greater when the linkage is based on 1,250 students in 50 schools, rather than merely on 50 school means. However, this is a quantitative issue, and for a state with an assessment that is highly correlated with NAEP, a linkage based on school means may be more precise than a linkage in another state, with an assessment less highly correlated with NAEP, based on individual student data. Since both the technical and political costs of creating a link between individual student NAEP and state assessment scores vary substantially

between states, the latter alternative was also evaluated (see Study Question #1).

Student Level Linkage. The most noticeable amount of effort required for the sample size reduction is for the creation of a linkage file that will allow an analyst to merge state assessment scores into the NAEP data file for the same students.

There are three distinct scenarios: (1) student level state assessment data are available to the State Education Agency and are filed with an identification code that can be entered on NAEP administration schedules at the time of NAEP sample selection; (2) student level state assessment assessment data are available to the State Education Agency but are filed only with the student and school names; and (3) student level state assessment data are not available to the State Education Agency.

Under the first scenario, the major effort is the data entry of the two identification codes, NAEP and state assessment, into a linkage file, for each of the students participating in NAEP. Under the second scenario, an additional step of looking up state assessment scores of the 25 NAEP-participating students in 50 schools by name is required. This procedure was used successfully in the ESSI study of NAEP-to-state assessment linkage. Under the third scenario, unless a particular exception can be made (e.g., because individual state assessment scores would not be linked to students' names on any file or report), a student level linkage may be impossible.

To preserve confidentiality, development of the linkage can be broken into separate steps. For example, at the time at which students are selected for participation in NAEP, a (spreadsheet) file can be created containing two numbers, the NAEP booklet identification code and a state assessment identification code, for each NAEP sampled student. That file can be split into two files, A and B, joined only by a common linkage code (for example, the row number in the spreadsheet). One of these files, A, which would contain only the state assessment identification code and the linkage code but no NAEP identification, could be merged with the state assessment data base when the scores become available (then dropping the state assessment identification code), to create a file containing only the state assessment scores and linkage code for each student on the file. That file could then be merged with the NAEP identification code, using the other linkage file, B, to create a file containing only the NAEP identification code and the state assessment scores for the student. That file is finally merged with the NAEP database to carry out the analyses required for reporting. These operations can be carried out in the state education assessment offices in a way that preserves individual confidentiality of both state assessment and the NAEP results.

School Level Linkage. Production of a file for a school level linkage is much simpler. School level state assessment data are made available as public information in most states, and in some cases, the data can be obtained merely by downloading a file from the State Education Agency's internet web page. The amount of effort for production of a school level assessment data and linkage file, once the state has summarized the data, is a matter of a few hours.

Step 4. Analyses.

The additional analyses required for using state assessment data to reduce the State NAEP sample size in a state are of three types: (1) verification analyses, (2) linkage development analyses, and (3) population estimation analyses. These will be carried out by the NAEP contractor.

Verification Analyses. It is not essential that the state assessment be administered in exactly the same way as NAEP is administered, but it is essential (for the validity of the linkage) that the state assessment be administered in schools participating in NAEP in the same way that the state assessment is administered in other schools. Any factors that might lead to lower state assessment scores in the NAEP participating schools than in other schools must therefore be checked. These verification analyses will be based on simple comparisons of statistics computed for the NAEP participating schools and for a matched set of schools selected by NAEP. They will include (1) verification that all schools in both half-samples participated in the state assessment, (2) a comparison of the percentages of student exclusions, (3) a comparison of absence rates, (4) a comparison of the ethnic patterns of exclusions and absence rates, and (5) a comparison of the mean and variance (and if the scores are multidimensional, the intercorrelations) of the state assessment scores. Although, due to sampling variability, these comparisons will undoubtedly show some differences, they should not be statistically significant differences -- that is, they should not be so large that they lead one to conclude that there was a systematic difference in the administration of the state assessment between the two samples.

Linkage Development Analyses. The linkage consists of a formula for estimating the mean and standard deviation of NAEP scores for every public school. Tests have indicated that such a formula can be developed using linear regression, either on school level data or on student level data. If student level data are used, school level variation (e.g., school mean state assessment scores) as well as individual level variation must be included in the formula, because in most states there is a significant component of between-school variation not accounted for merely by variation among students. A random component, normally distributed and with the standard deviation given by the regression, is added to each school mean, so that the resulting distribution of school means matches both the mean and the standard deviation of the NAEP scores they are estimating.

Estimation of NAEP results. The overall NAEP mean for the state is estimated by averaging together the NAEP school means for schools participating in NAEP and the estimated NAEP means, based on the linkage to state assessment scores, for all other public schools in the sampling frame. For population subgroups that vary within schools (e.g., race/ethnicity), the analysis depends on whether the population subgroup distributions are known for schools not participating in NAEP. If known, then within-school differences can be predicted (from the differences within NAEP schools) and used to produce precise subpopulation estimates. If not known for other schools. then reports of the subpopulation differences in the state would be based solely on the NAEP schools, resulting in larger standard errors, by a factor of about 1.5.

These analyses are in addition to (a) the NAEP scaling analyses that are to be carried out

in any case and (b) state assessment analyses that are to be carried out in any case. The timing of the state assessment is important for this method, however, because NAEP cannot afford to be delayed due to the failure to receive state assessment data at the time that the NAEP scaling analyses are completed. This should not be a problem, however, even though state assessments are administered later in the school year than State NAEP. States generally require their assessment contractor to produce reports of the results of the assessment in a time frame either shorter or equivalent to the time it takes for NAEP to carry out analyses (i.e., by September, following the February administration).

Step 5. Reporting.

Reporting will be the same as in the past. However, if the analyses resulted in rejecting the linkage, the published results would be based on half as many cases as in the past. This would mean that standard errors would be greater and that reporting for some subgroups in the state might need to be suppressed. On the other hand, if the linkage is acceptable and is stronger than planned for (i.e., NAEP is more highly correlated with the state assessment than was planned for), the standard errors will be correspondingly smaller and reports on subgroups more precise.

It should be noted that NAEP collects not only cognitive achievement information but also student, school, teacher, and classroom background information. This additional information provides the basis for reporting not only the overall distribution of NAEP performance in a state but also the performance of subpopulations of students defined by the background information. Some of this information (e.g., school size and percentage of minority enrollment) is available from the Common Core of Data in essentially equivalent form, but unless the state collects the other background information on schools not participating in State NAEP, reports relating performance to these background measures will be based on the half-sample and therefore less precise. The standard error of statistics will be roughly 1.5 (i.e., close to the square root of 2) times as large as they would be if based on a full sample.

Study Questions

The overall question, whether state assessment data (through linkage with NAEP) can be used to reduce state NAEP sample sizes, can be broken into 4 different aspects:

Question #1: How will sample size reduction using state assessment data affect accuracy of State NAEP?

Question #2: What will the costs be?

Question #3: How will confidentiality requirements be fulfilled?

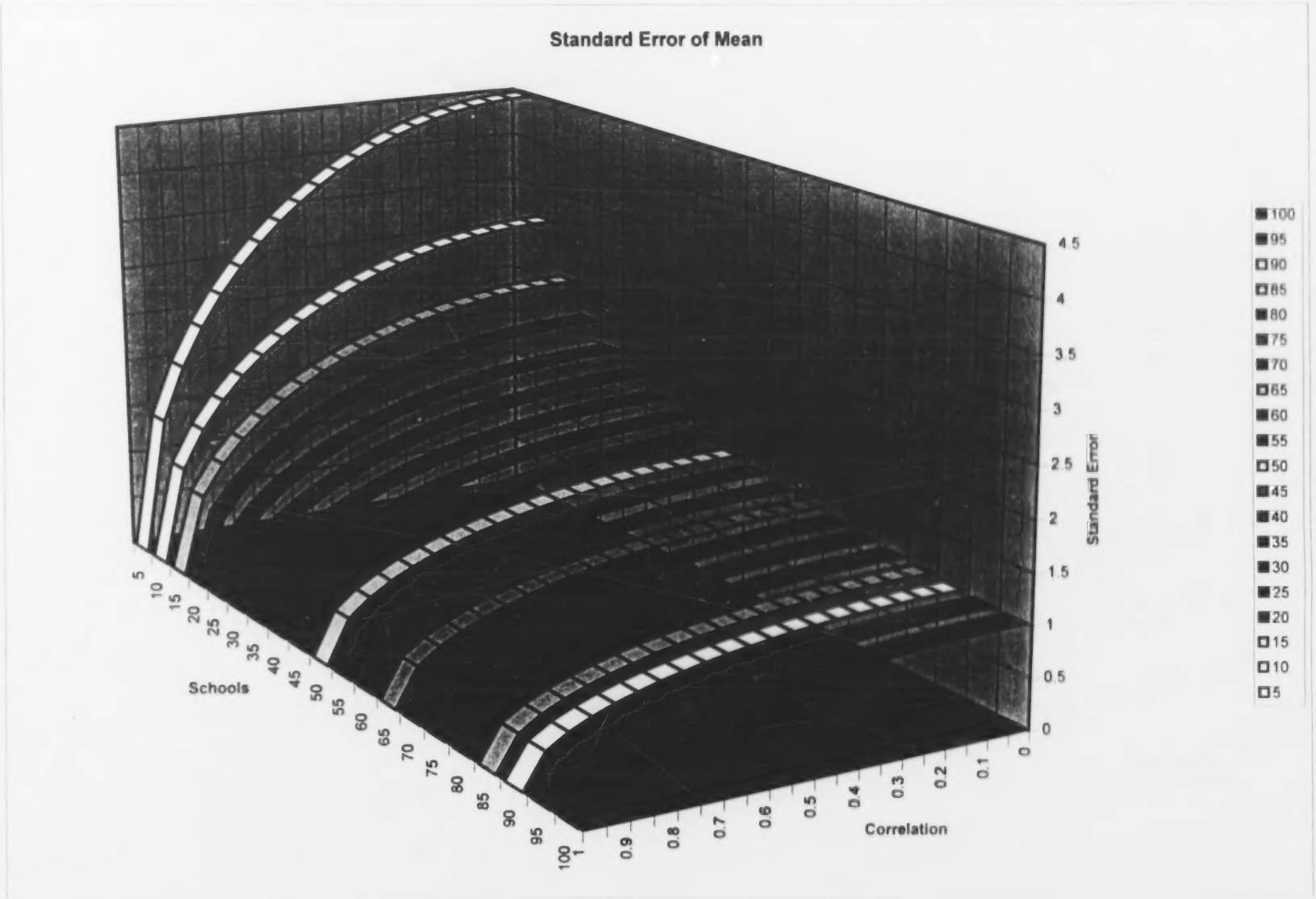
Question #4: How many states can be expected to participate?

Overarching these questions is a crucial issue -- *Are student level linkage data needed, or is a school level linkage database sufficient?* This question is crucial, because relying on school-level data alone (1) reduces the costs, which are already small, dramatically, by eliminating the need for data entry for a student level linkage database, (2) eliminates any issues of individual student score confidentiality, and therefore, (3) can be expected to facilitate the participation of a larger set of states. Therefore, part of the work to address Question #1 was to evaluate the relative precision of estimates based on school-level and student-level linkages.

Question #1: How will sample size reduction using state assessment data affect accuracy of State NAEP?

In a sense, State NAEP consists of 40 to 50 separate assessments, all using the same instruments and data collection and analysis methods. The accuracy of results in one state have only an indirect effect on the outcomes of the assessment in other states. Therefore, for the most part, this question can be addressed at the state level. The question concerns the trade-off between a database of 100 schools in a state participating in NAEP, versus a linked database of 50 schools participating in NAEP plus all of the schools in the state participating in a correlated assessment. In states in which there are substantially more than 100 schools serving a particular grade and in which there is a moderately high correlation between NAEP and the state assessment, the trade-off favors the linked database. The relation between number of schools participating in NAEP, correlation between assessments, and the standard error of NAEP estimates is shown in Figure 1, where the standard error of 1.0 is arbitrarily set at the current implementation of 100 schools and no state data (i.e., a correlation of 0). It can be seen in Figure 1 that the reduction in standard errors at high correlations can easily more than offset the increase in standard errors in the range of 50 to 100 schools.

The reason this trade-off works is that, although error is added by basing estimates for non-NAEP schools on an imperfect statistical linkage, that error is smaller than the sampling


Standard Error of Mean
Schools
Correlation
Standard Error
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5

error, which is eliminated if estimates are obtained for all schools in the state. The approximate specific formula for the relationship between correlation and sample size is:

$$1 - r^2 = n/100$$

That is, to obtain roughly the same accuracy as obtained using 100 schools and no state assessment information, if the correlation between NAEP and a state assessment is .7, then only 51 schools are required ((1 - .49) = 51/100). Although this is approximate, because differential weighting of scores and differential contributions of within- and between-school variation are not taken into account, it is borne out in model simulations for correlated assessments.

Are student-level state assessment data needed for the linkage?

To address the question of how much loss of precision would occur if the linkage were based solely on school-level information, the data from the 1996 State NAEP in mathematics in three states, plus state assessment scores for the same students, were used. With these data, the precision of the regression based on one half the NAEP sample for predicting the sample mean in the other half of the sample could be directly estimated by repeated sampling of halves of the data. (To take between-school variation into account, the half-samples were samples of schools.) Simulation with actual NAEP and state assessment data is essential for this comparison, because the outcome depends on the extent to which variation in NAEP performance and in state assessment performance is between- or within-schools. Student-level data contribute more to the estimation if most of the variance in performance is between students in the same schools.

For this estimation, the model used was the simple linear model:

$$NAEP_{ij} = a + b_1 x\, State_{ij} + b_2 x\, State_{i}$$

where the NAEP measure was the mean of five composite plausible values for participant j in school i, and where both the state assessment score for the individual (i j) and the school mean state assessment score (i .) entered into the equation. (For this simulation, the mean of NAEP participants in the school was used for the school mean.) For the corresponding school-level model, the regression treated schools as observations. The statistic compared is the standard error of the estimated NAEP mean for the half-sample not used in the estimation, as measured by the standard deviation of values in repeated random half-samples.

The results are presented in Table 1. The tabulated values are the ratios of the average standard error using a state-assessment-based estimate to the average standard error using the actual NAEP data for the same schools. The values for each of three states and two grades are based on 100 random half samples. Two sets of values are shown, (1) using the simple regression estimates, and (2) imputations which augment the error variance of the regression estimates to match the standard deviation of NAEP scores. As can be seen in Table 1, for the

simple regression estimates, there is little loss in precision from limiting the database to school-level data. The standard errors of means based on linkage are generally in the same range as the actual standard errors for the same schools (estimated by repeated half-sample variation). In practice, however, the imputed values would be used, in order to avoid underestimation of population variances. In this case, the school-level standard errors are about one-sixth larger (1.26/1.08) than the student-level standard errors. This ratio corresponds to the difference in r squared between .6 and .7. That is, if one were to decide that a value of .6 for r squared were adequate for the linkage using student-level data, then setting a requirement for an r squared of .7 would be reasonable for use of school-level data.

Table 1. Ratio of regression-based standard errors to actual standard errors for the same (half-)sample, using school-level or student-level data in the linkage.

		Imputed Estimate of Mean		Regression Estimate of Mean	
State	Grade	School-level	Student-level	School-level	Student-level
1	4	1.18	1.14	0.91	1.04
	8	1.36	1.20	1.08	1.19
2	4	1.29	1.11	1.02	0.99
	8	1.19	1.16	1.06	1.06
3	4	1.27	0.89	1.00	0.88
	8	1.25	0.95	1.05	0.92
Total		1.26	1.08	1.02	1.01

Note: The values are generally greater than 1.00 because the NAEP and state assessment samples contained the same number of schools.

Similar results can be expected to hold for percentages of students scoring above specified cutpoints, provided both means and standard deviations of school score distributions are modeled. For population subgroup statistics for which state assessment data can be disaggregated statewide, the same argument holds. For population subgroups for which only school distributions are known (e.g., the percentage of minority students in the school), differences between scores for subgroups must be based on the reduced sample, yielding somewhat larger standard errors (depending on the stability of the differences across schools).

To summarize, if the state assessment has a squared correlation of .7 (or higher) with NAEP, and if the NAEP sample is a subset of the public schools in the state for which there is state assessment information, the precision of overall mean estimates would be improved by using state assessment data, even if the NAEP sample were cut from 100 to 50. If, after the data

BEST COPY AVAILABLE

are collected, analyses indicate that the linkage is not possible, the results would be that standard errors for all NAEP statistics for that state would be increased by a factor of about 1.5. In any case, for subpopulation charts for which no disaggregated state assessment data are available, there will be an increase in standard errors by a factor of about 1.5.

Question #2: What will the costs be?

Reducing the sample size of State NAEP in a state will, of course, save printing, test administrator training, data collection, and scoring costs. In order not to pay for those reduced costs by reducing precision, state assessment data are to be linked to NAEP to enhance the precision of estimation. The amount of effort required for the additional activities involved inthe linkage is estimatable, based on activities in current projects to link NAEP to state assessments.

The costs can best be estimated in terms of the number of (additional) professional hours required, for each step in the process, for each state. The five steps are invitation, sample selection, linkage data file production, analysis, and reporting. The hours indicated are conservative (high) estimates for the amount of effort required in addition to NAEP activities that would be carried out even if the state was not participating in the sample size reduction initiative.

Invitation. First, analyses need to be done to decide whether it is appropriate to invite each state participating in NAEP to take advantage of the sample size reduction. Acquisition of school means and computation of correlations might take as much as 8 hours per state. Second, a conversation with the state testing director should take place, so that information is available about the state assessment -- e.g., how has it changed this year? when can the results be obtained? what special confidentiality restrictions exist? This might require as much as 4 hours per state. Total: 12 hours per state.

Sample Selection. The only additional activity involved in selecting the sample of schools is the retention of the names of half of the schools, for later use as a state assessment verification sample. Total: 2 hours per state.

Linkage File Production. If school level data are to be used, all that is needed is the recording of state score distributions (means and standard deviations) for each of 50 NAEP schools. This might take as much as 8 hours, if data cannot be retrieved directly via the internet. If student level data are to be used, then a process of recording identification codes linking NAEP to state assessment scores must be carried out. For 30 participants at each of 50 schools, this process can take a total effort of as much as two weeks of training, look-up, data entry, and checking, or 80 hours. Total: 8 or 80 hours per state.

Analysis. Arranging for the acquisition of the state assessment file and transforming the state assessment data to merge with NAEP might require as much as 16 hours. Although the programming for the analyses will have been completed once for all states, the three analytical

steps, verification, linkage parameter estimation, and population estimation, might take as much as an additional 16 hours. Total: 32 hours per state.

Reporting. A small amount of effort might be required for adapting a description of the procedures used to use in the state, and (parts of) one or two meetings with press and/or constituencies might be necessary to explain the fact that accuracy was not sacrificed in using the state assessment to enhance NAEP estimates on a smaller sample. Total: 16 hours.

The total amount of effort estimated is either 70 or 142 hours per state, depending on whether a student-level linkage is needed. In my opinion, the amount of effort will be much less in most states; however, with this estimate, allowances can be made for additional activities to deal with special needs in some states. Depending on labor costs, the effort is in the range of $5,000 - $10,000 per state.

Question #3: How will confidentiality requirements be fulfilled?

If a student-level linkage is required, it must be developed in a manner that conforms to the confidentiality assurances given for both State NAEP and the state assessment. Methods for assuring that no release of student level data can occur were described (above) in spelling out the plan (see Step #3). However, if state legal requirements prohibit the use of state assessment data for this purpose, no amount of caution will be sufficient to enable the linkage to be produced. In those states, a student-level linkage is impossible.

Generally, there will not be a significant problem with school-level scores. However, the file development and analyses must be undertaken in such a way as to keep the identification of particular NAEP schools from becoming public. Therefore, the files associating state assessment scores with NAEP data must be considered restricted. Also, in states where the only level of assessment data available for this purpose is the district level, the linkage may be difficult.

Question #4: How many states can be expected to participate?

Of 43 states participating in State NAEP in mathematics in 1996, examination of information published by the Council of Chief State School Officers[2] indicates that about 31 had state assessments that might support the linkage necessary for the NAEP sample size reduction plan. Some other states also had state assessment programs, but the information provided by those were either at a grade level not relevant to State NAEP (e.g., sixth or ninth grade) or were of a mastery nature that precluded use in predicting variation in NAEP performance.

---

[2] Roeber, E., Bond, L., & Braskamp, D. *Annual Survey of State Student Assessment Programs: Fall 1996*. Council of Chief State School Officers, Washington, 1997.

Also, of the 43 states, 11 were "asterisked" in the NAEP reports for failing to reach the highest school participation rate requirements. These states, plus at least one or two that were not able to participate, would be very likely to participate if there were a chance for school sample size reduction. Of course, some of the states finding NAEP participation most difficult also do not have state assessments with the characteristics needed for this sample size reduction plan. Furthermore, some of the smallest states have so few schools that it might be unworkable to cut the number of schools in half.

Inevitably, the characteristics of state assessments change from year to year. Therefore, it is impossible to predict with any accuracy the number of states that might participate in the sample size reduction plan in a particular year. However, if it were shown to work, probably at least half of the states would be able to and would choose to reduce NAEP school sample sizes.

Conclusions

State assessment data can be used in many states to increase the precision of State NAEP, to the point that the current level of precision for overall state population performance estimates can be maintained with samples of half as many schools. While the linkage of student-level assessment scores is ideal, linkages based solely on school-level summary statistics appear to be sufficient, when the correlation between tests is high. Many states appear to have assessments with such correlations to NAEP.

Creation of student-level linkages while maintaining confidentiality of student scores is feasible in many states, but in others restrictions on the use of state assessment data may preclude development of a student-level linkage. Restrictions in each state should be explored to determine customized alternatives, such as carrying out different steps in the analysis at different sites.

The cost of the linking procedures required for implementing the sample size reduction is in the range from $5,000 - $10,000 per state, which is a small percentage of the cost of the administration of State NAEP in an additional 50 schools in the state. The state and local portion of that latter cost was estimated by the National Academy of Education to be about $25,000 in 1990.

Many states have difficulty recruiting schools to participate in NAEP and would welcome this initiative. In fact, they appear willing to take the gamble that their state assessment would support the plan, realizing that their State NAEP reports might have somewhat larger standard errors if the linkage proved, after the fact, to be insufficient. In those cases, the reports would be based solely on the half-sample of schools that participated in NAEP.

The analyses carried out here suggest that using a school-level linkage, several states

could be involved in sample size reduction, possibly as early as 1998. Analyses that AIR has carried out using the 1994 reading assessment data indicate that performance data on many state reading assessments are correlated with NAEP reading performance.

Although the NAEP redesign does not necessarily include the 1998 State NAEP in reading and writing, I recommend exploring this with the states that have potentially useful state assessment data. On a research basis, this might be tried in a handful of states, at one grade level, say grade 8. The results of such a pilot study would provide information on both the advantages and disadvantages of the use of state assessment data to reduce NAEP sample sizes.

The State NAEP sampling design overlaps assessments, such as reading and writing, in the same schools. Therefore, to be effective, the sample size reduction needs to be applied to both NAEP assessments being administered in the same schools. Although there is ample data to indicate that NAEP reading and mathematics scales are correlated with state assessment data, this evidence has not been developed for NAEP writing and science assessments. Nevertheless, because NAEP reading and mathematics assessments are found to correlate with state assessments in science and language arts, as well as mathematics and reading, there is a reasonable expectation that linkages will be possible in the other areas as well.

Listing of NCES Working Papers to Date

Please contact Ruth R. Harris at (202) 219-1831
if you are interested in any of the following papers

Number	Title	Contact
94-01 (July)	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94-02 (July)	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94-03 (July)	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94-04 (July)	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94-05 (July)	Cost-of-Education Differentials Across the States	William Fowler
94-06 (July)	Six Papers on Teachers from the 1990-91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
94-07 (Nov.)	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
95-01 (Jan.)	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95-02 (Jan.)	QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95-03 (Jan.)	Schools and Staffing Survey: 1990-91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk
95-04 (Jan.)	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95-05 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
95-06 (Jan.)	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95-07 (Jan.)	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95-08 (Feb.)	CCD Adjustment to the 1990-91 SASS: A Comparison of Estimates	Dan Kasprzyk
95-09 (Feb.)	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95-10 (Feb.)	The Results of the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95-11 (Mar.)	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95-12 (Mar.)	Rural Education Data User's Guide	Samuel Peng
95-13 (Mar.)	Assessing Students with Disabilities and Limited English Proficiency	James Houser
95-14 (Mar.)	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95-15 (Apr.)	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95-16 (Apr.)	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95-17 (May)	Estimates of Expenditures for Private K-12 Schools	Stephen Broughman
95-18 (Nov.)	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96-01 (Jan.)	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-02 (Feb.)	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96-03 (Feb.)	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
96-04 (Feb.)	Census Mapping Project/School District Data Book	Tai Phan
96-05 (Feb.)	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96-06 (Mar.)	The Schools and Staffing Survey (SASS) for 1998-99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96-07 (Mar.)	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96-08 (Apr.)	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96-09 (Apr.)	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998-99 SASS	Dan Kasprzyk
96-10 (Apr.)	1998-99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96-11 (June)	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96-12 (June)	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96-13 (June)	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96-14 (June)	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
96-15 (June)	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96-16 (June)	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96-17 (July)	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
96-18 (Aug.)	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
96-19 (Oct.)	Assessment and Analysis of School-Level Expenditures	William Fowler
96-20 (Oct.)	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96-21 (Oct.)	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96-22 (Oct.)	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
96-23 (Oct.)	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96-24 (Oct.)	National Assessments of Teacher Quality	Dan Kasprzyk
96-25 (Oct.)	Measures of Inservice Professional Development: Suggested Items for the 1998-1999 Schools and Staffing Survey	Dan Kasprzyk
96-26 (Nov.)	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96-27 (Nov.)	Intersurvey Consistency in NCES Private School Surveys for 1993-94	Steven Kaufman

Listing of NCES Working Papers to Date--Continued

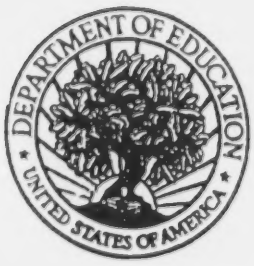
Number	Title	Contact
96-28 (Nov.)	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
96-29 (Nov.)	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96-30 (Dec.)	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-01 (Feb.)	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk
97-02 (Feb.)	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-03 (Feb.)	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97-04 (Feb.)	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-05 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97-06 (Feb.)	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97-07 (Mar.)	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97-08 (Mar.)	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-09 (Apr.)	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97-10 (Apr.)	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993-94 School Year	Dan Kasprzyk
97-11 (Apr.)	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97-12 (Apr.)	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97-13 (Apr.)	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
97-14 (Apr.)	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97-15 (May)	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97-16 (May)	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns
97-17 (May)	International Education Expenditure Comparability Study: Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns
97-18 (June)	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97-19 (June)	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe
97-20 (June)	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97-21 (June)	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
97-22 (July)	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman

Listing of NCES Working Papers to Date--Continued

Number	Title	Contact
97-23 (July)	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97-24 (Aug.)	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97-25 (Aug.)	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
97-26 (Oct.)	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
97-27 (Oct.)	Pilot Test of IPEDS Finance Survey	Peter Stowe
97-28 (Oct.)	Comparison of Estimates in the 1996 National Household Education Survey	Kathryn Chandler
97-29 (Oct.)	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Steven Gorman


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA